Identification of the sequences recognized by phage φ29 transcriptional activator: possible interaction between the activator and the RNA polymerase

Beatriz Nuez, Fernando Rojo, Isabel Barthelemy+ and Margarita Salas*
Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

Received January 31, 1991; Revised and Accepted April 5, 1991

ABSTRACT

Expression of *Bacillus subtilis* phage φ29 late genes requires the transcriptional activator protein p4. This activator binds to a region of the late A3 promoter spanning nucleotides −56 to −102 relative to the transcription start site, generating a strong bendingTin the DNA. In this work the target sequences recognized by protein p4 in the phage φ29 late A3 promoter have been characterized. The binding of protein p4 to derivatives of the late A3 promoter harbouring deletions in the protein p4 binding site has been studied. When protein p4 recognition sequences were altered, the activator could only bind to the promoter in the presence of RNA polymerase. This strong cooperativity in the binding of protein p4 and RNA polymerase to the promoter suggests the presence of direct protein-protein contacts between them.

INTRODUCTION

The prevalent model for the process of transcription activation by prokaryotic regulatory proteins involves at least two steps: the activator first binds to DNA in the vicinity of the promoter and then facilitates binding of the RNA polymerase and/or the formation of a transcriptionally active open complex. Transcription activators frequently bind to DNA as dimers recognizing symmetric binding sites consisting of short inverted repeats. This seems to be a convenient way to increase the interactions with DNA without the need of synthesizing excessively large proteins (16). Bendability of the binding site can also play a significant role in the binding process in those cases where the DNA accommodates or wraps around the protein (21). In several promoters, the activation process appears to be induced by direct contacts between the activator and the RNA polymerase (reviewed in 1) although in some systems in which the activator induces a significant increase in DNA curvature, this change in promoter geometry also seems to play a role in the activation mechanism (3,14). Local conformational changes in the structure of DNA, realigning the −10 and −35 regions of the promoter, have been proposed to mediate activation of the Tn*501* mercury resistance operon by the MerR-Hg(II) activator (12). Nevertheless, only a few activators have been studied in detail and their activation mechanisms are still far from being completely understood.

Transcription of *B. subtilis* phage φ29 late genes occurs from the A3 promoter (P_{A3}), located at a region in which other early promoters are also present (Fig. 1). The viral early protein p4 has been shown to be a transcriptional activator, required for P_{A3} transcription (18,2). By DNase I footprinting, it has been shown that protein p4 binds to a region of the promoter located between nucleotides −56 to −102 relative to the transcription start site. This region contains two 8 bp inverted sequences which have been proposed to be protein p4 recognition sites (2). In addition, this region of the promoter has a sequence-directed curvature that increases considerably when protein p4 binds to it (2,14). The behaviour of a number of carboxyl-terminal mutants of protein p4 has allowed us to propose a model for its interaction with the A3 promoter in which the curvature is induced in two steps: first, binding of two monomers of protein p4 to the inverted sequences and subsequent interaction between them would generate a bend between the binding sequences; second, the highly basic carboxyl terminus of the protein would establish non-specific electrostatic interactions with the negatively charged DNA backbone, inducing a bend at both ends of protein p4 binding region. The complete induction of this curvature appears to be required to favour the binding of *B. subtilis* σ^A-RNA polymerase to the promoter (14). This suggests that the curvature induced by protein p4 has indeed a role in the process of transcription activation. In principle, curvature could provide a DNA conformation suitable for an efficient RNA polymerase binding or could simply correctly orient protein p4 so that it can adequately interact with RNA polymerase. In this latter case, protein-protein contacts between the activator and RNA polymerase would be important in the activation mechanism. The two possibilities are, however, not mutually exclusive.

* To whom correspondence should be addressed

+ Present address: Antibióticos Farma S.A., Antonio Lopez 109-111, 28026 Madrid, Spain

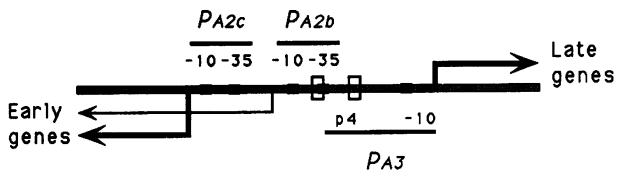

Figure 1. Transcriptional map of the phage ϕ29 late A3 promoter region. Black boxes indicate the −10 and −35 regions of each promoter and open boxes the two inverted sequences recognized by protein p4 within the A3 promoter. The early A2b and A2c promoters are also shown and their transcriptional start sites are marked by arrows. The thickness of the arrows indicates the in vitro strength of each promoter (11).

In this paper we describe the characterization of the protein p4 recognition sequences at the late A3 promoter and present evidence suggesting that the activator and the RNA polymerase directly contact each other.

METHODS

DNA manipulations

Plasmid DNA was purified using standard methods (10). ϕ29 DNA was prepared as described (8), and appropriate restriction fragments containing the A3 promoter were purified from polyacrylamide gels by overnight diffusion at 37°C in 0.5 M ammonium acetate, 0.1% SDS, 1 mM EDTA, or from agarose gels using DEAE-nitrocellulose membranes from Schleicher and Schuell. DNA fragments were labelled at their 3′ ends with Klenow enzyme or at their 5′ ends with T4-polynucleotide kinase. Sequencing reactions used as DNA size markers were performed by chemical cleavage (9).

Missing nucleoside assay

A 198 bp AccI-HindIII DNA fragment from bacteriophage ϕ29 containing the A3 promoter, spanning from position 5063 to 5261 of ϕ29 genome (23), was labelled at either strand and treated with hydroxyl radicals as described (7), to obtain DNA fragments which statistically had no more than one gap per molecule. DNA was precipitated and resuspended in 20 μl of a solution containing 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$ and 2.5 μg of poly (dI-dC). Protein p4 (1 μg) was then added and the mixture incubated for 30 min at 4°C. Bound and unbound DNA were separated in a 4% polyacrylamide gel in 12 mM Tris-acetate (pH 7.5), 1 mM EDTA, purified by electroelution and analyzed in denaturing polyacrylamide gels.

Construction of phage ϕ29 A3 promoter deletion mutants

The ϕ29 AccI-HindIII restriction fragment mentioned in the preceding section, containing the A3 promoter, was filled-in with Klenow enzyme and cloned in the SmaI site of *M13mp19*. The recombinant obtained, containing P_{A3}, was named *mA3*. A collection of A3 promoter deletion mutants lacking an increasing number of base pairs from the phage ϕ29 AccI site (coordinate 5063 of the ϕ29 genome) was obtained by Bal 31 digestion of *mA3* replicative form linearized with EcoRI. Therefore, deletions proceeded from ϕ29 position −124 relative to the P_{A3} transcription start site towards the protein p4 binding region, also affecting *M13* sequences upstream of the EcoRI site (position 6287). Three different derivatives were characterized which conserved the A3 promoter up to positions −95, −88 and −87

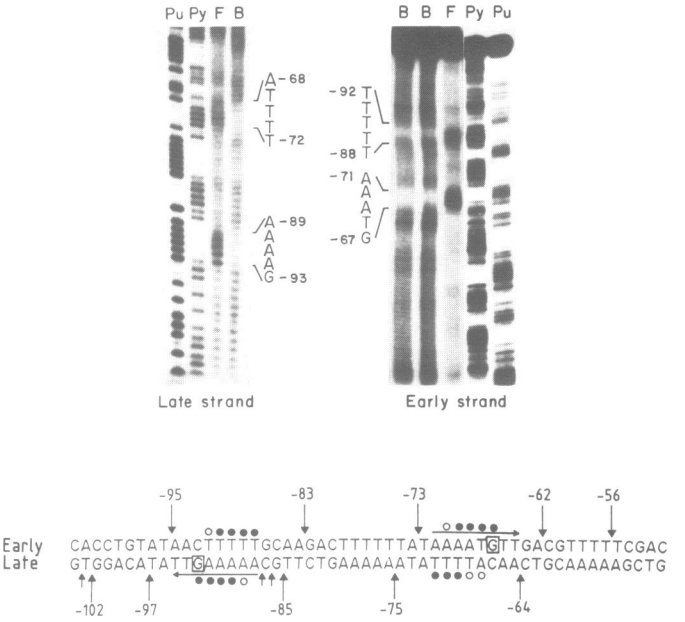

Figure 2. Missing nucleoside assay of the binding of protein p4 to the A3 promoter. Protein-bound (B, shown in duplicate for the early strand) and free (F) DNA were resolved in a 4% polyacrylamide gel, purified by electroelution, and analyzed in 6% (late strand) or 8% (early strand) denaturing polyacrylamide gels. Early and late strands correspond to the template strand for the early and late genes, respectively. Nucleosides whose absence interferes with protein p4 binding are indicated for both strands. Pu and Py are Maxam and Gilbert purine and pyrimidine sequencing reactions carried as size markers. Numbers indicate positions relative to the transcription start site. The sequence of the protein p4 binding region is shown at the bottom. The inverted sequences are indicated by arrows and the nucleosides critical for protein p4 binding, by filled circles. The open circles correspond to nucleosides whose absence diminishes but does not totally impair protein p4 binding. Guanine residues whose methylation interferes with protein p4 binding are enclosed in rectangles and DNase I hypersensitive sites indicated by vertical arrows (2).

relative to the transcription start site (nucleotides 5093, 5100 and 5101 from the left end of ϕ29 DNA) which lie adjacent to the *M13* positions 6318, 6323 and 6296, respectively. They were named *mA3Δ95*, *mA3Δ88* and *mA3Δ87*, and contained promoters P_{A3}Δ95, P_{A3}Δ88 and P_{A3}Δ87.

In vitro transcription assay

The DNA fragments used as templates for the *in vitro* transcription assays were obtained by digestion of *M13* derivatives replicative forms with HindIII and EcoRI (*mA3*, wild-type) or HindIII and PvuI (*mA3* deletion constructs). The sizes of the fragments obtained were 250 bp (P_{A3} wild-type), 299 bp (P_{A3}Δ95), 286 bp (P_{A3}Δ88) and 313 bp (P_{A3}Δ87). Reactions were done in 30 mM triethanolamine-HCl (pH 7.9), 8 mM $MgAc_2$, 4% glycerol, 96 mM ammonium sulfate, 0.2 mM each ATP, CTP and GTP, 80 μM [α-^{32}P]UTP (2 μCi), in a total volume of 25 μl, and contained 1 μg of protein p4, purified as described (14), 0.25 μg of purified *B. subtilis* σ^A-RNA polymerase (18) and the indicated amounts of template DNA. Reactions were carried out for 10 minutes at 37°C and stopped with 75 μl of a solution containing 30 mM triethanolamine-HCl (pH 7.9), 8 mM $MgAc_2$, 0.15 mg/ml tRNA and 0.15% SDS. Samples were phenol extracted, filtered through a 1 ml Sephadex G50 spun column, ethanol precipitated and analyzed in a 4% denaturing polyacrylamide gel. After autoradiography and

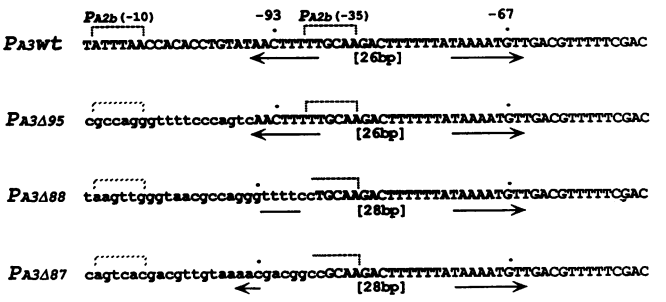

Figure 3. Sequence of the A3 promoter and deletion mutants. Phage ϕ29 sequences are in capital letters whereas vector sequences are in low case. Inverted sequences recognized by protein p4 are indicated by arrows, and positions −67 and −93 relative to the transcriptional start site are shown. The sequences underlined in the deletion derivatives show regions of residual homology with the wild-type distal inverted sequence. Numbers in brackets indicate the distance between the inverted sequences, estimated between guanine residue −67 and either cytosine residue −93 (wild-type and A3Δ95 promoters) or that considered to be homologous in the deletion derivative.

densitometry, protein p4 stimulation was calculated as the ratio between the amount of transcript from the late promoter obtained with and without protein p4.

DNase I footprinting

Binding of protein p4 and/or *B. subtilis* σ^A-RNA polymerase to DNA fragments containing phage ϕ29 A3 promoter, or its deletion derivatives, was analyzed by DNase I footprinting (5), using the same DNA fragments as in the transcription assays, labelled at either of their ends. Binding reactions were carried out in a final volume of 20 μl, in 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$ in the presence of 2 μg of poly (dI-dC) as non-specific competitor DNA; when indicated, 1μg of protein p4 and/or 0.5 μg of *B. subtilis* σ^A-RNA polymerase were added. After incubation for 15 min at 37°C, samples were transferred to room temperature and 50 ng of DNase I were added. Digestion was allowed to occur for 2 min and stopped with 1 μl of 0.5 M EDTA. DNA was precipitated and analyzed in denaturing polyacrylamide gels.

RESULTS

Determination of protein p4 recognition sequences at the ϕ29 late A3 promoter

Previous studies (2) have indicated that protein p4 binds to a region of the A3 promoter spanning from nt −56 to −102 relative to the transcription start site. This region contains two 8 bp long inverted sequences that were proposed to be protein p4 recognition sites (2). To analyze the importance of each nucleoside in the binding of protein p4 to P_{A3}, the missing nucleoside technique (7) was used. Random gaps were generated in the DNA by treatment with hydroxyl radicals and protein p4 was then allowed to bind to the gapped DNA. Bound and unbound DNA were separated in a band-shift gel, purified and analyzed in denaturing polyacrylamide gels. Fig. 2 shows that the digestion pattern of free DNA was enriched in five of the eight bands corresponding to each of the two inverted sequences proposed to be the protein p4 recognition target. The same bands were greatly decreased in the pattern of p4-bound DNA, implying that they are required for protein p4 binding and demonstrating the importance of both inverted sequences for the binding of protein p4 to DNA.

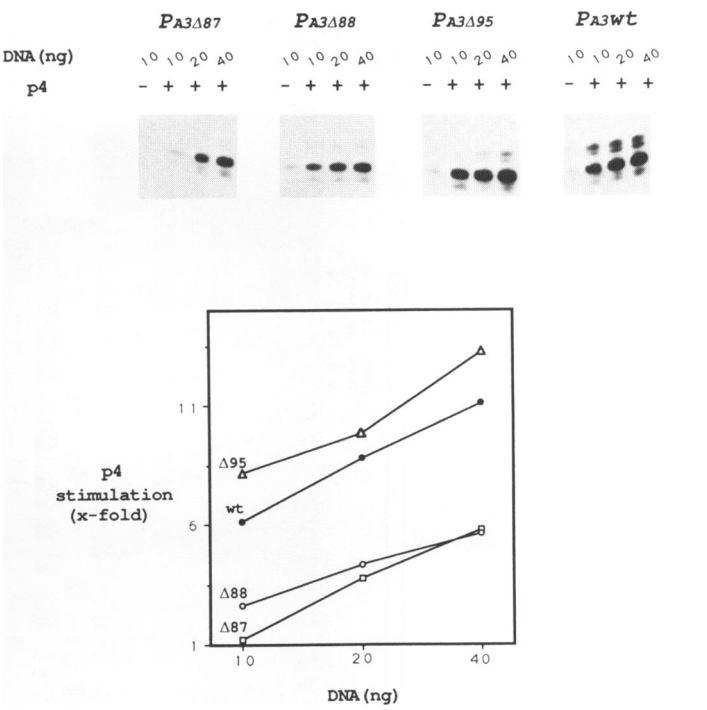

Figure 4. Activity of the A3 promoter and its deletion derivatives. Run-off transcripts produced from the A3 promoter, detected by autoradiography of the corresponding denaturing polyacrylamide gels, were quantified by densitometry; the graphic shows the ratio between the amount of transcript obtained in the presence and in the absence of protein p4 for different amounts of DNA template.

Activity of the A3 promoter deletion mutants

Using nuclease Bal 31, three derivatives of P_{A3} were obtained, named $P_{A3}\Delta95$, $P_{A3}\Delta88$ and $P_{A3}\Delta87$, harbouring deletions in protein p4 binding region upstream of positions −95, −88 or −87, respectively. The extent of each deletion is shown in Fig. 3. The deletion in mutant $P_{A3}\Delta95$ did not affect protein p4 recognition sites, but mutants $P_{A3}\Delta88$ and $P_{A3}\Delta87$ had lost part of these sequences, particularly those located in the upstream part of the inverted repeat. The ability of protein p4 to activate transcription from the P_{A3} deletion mutants was analysed by *in vitro* run-off assays. All mutants produced transcripts of identical size and, as shown in Fig. 4, an increase in the amount of template DNA up to 40 ng was correlated with an increase in the protein p4-dependent transcription of all the templates. The activity of mutant $P_{A3}\Delta95$, which conserves the two inverted sequences that constitute the complete protein p4 recognition site, was similar to that of the wild-type promoter, whereas mutants $P_{A3}\Delta87$ and $P_{A3}\Delta88$, which have the distal inverted sequence substituted by vector sequences, showed about 5 and 2.3 times less activity, respectively, than the wild-type P_{A3} when assayed at low DNA concentration. The residual activity in the mutants could be accounted for by some partial homology of the vector sequences with the deleted sequences (Fig 3). Alternatively, it could reflect that protein p4 is still able to recognize its binding site, although with low efficiency, when only half of its recognition sequences are maintained. This possibility was investigated by DNase I footprinting.

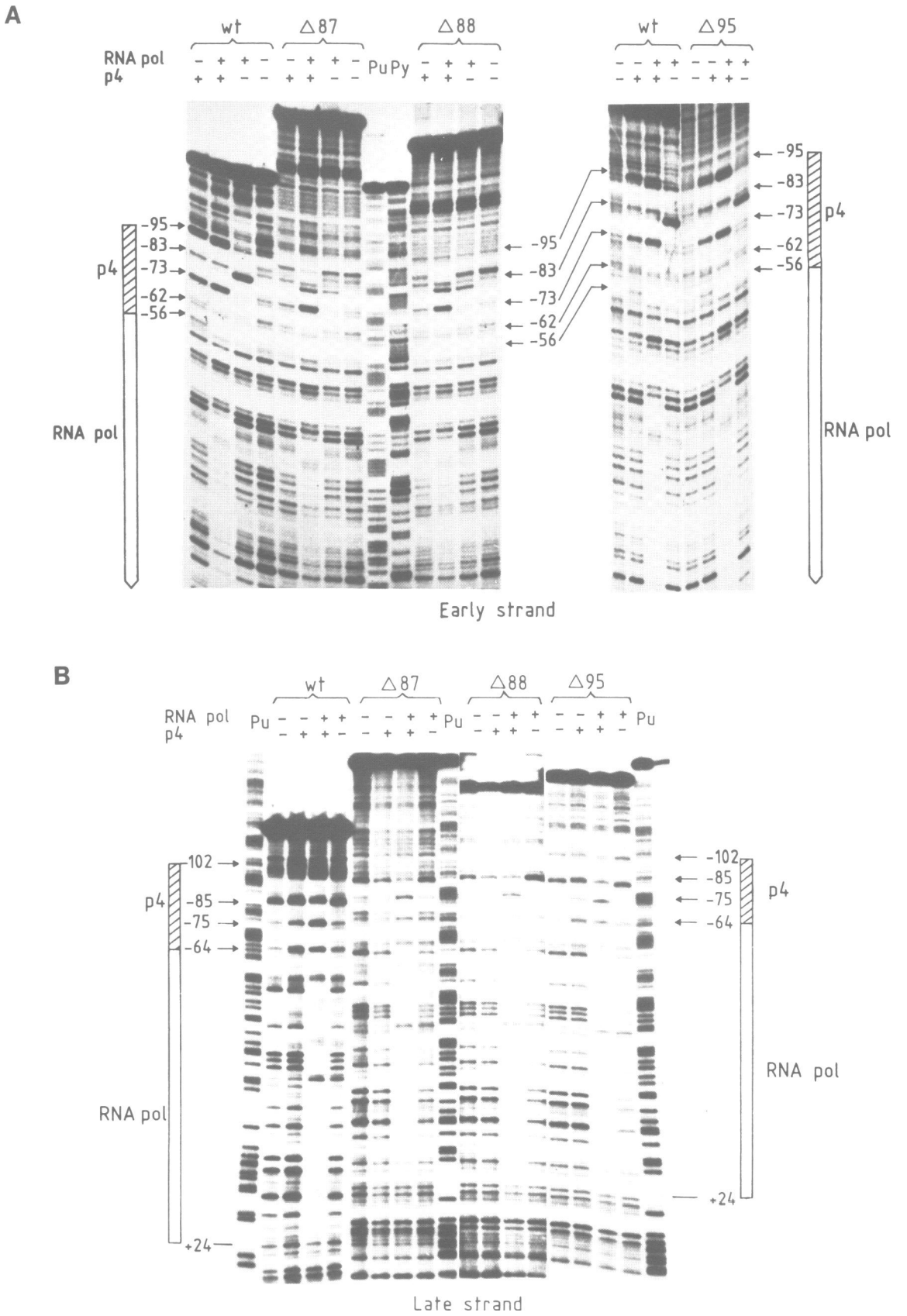

Figure 5. DNase I footprinting of protein p4 and RNA polymerase bound to the ϕ29 A3 promoter and its deletion derivatives. Positions that become hypersensitive to DNase I attack upon protein p4 binding are indicated by arrows. The binding regions are shown in rectangles, stripped for protein p4 and grey for RNA polymerase. (A) early strand; (B) late strand. Pu and Py are Maxam and Gilbert sequencing reactions carried as size markers.

Binding of protein p4 and RNA polymerase to the A3 promoter deletion mutants

Binding of protein p4 to the A3 promoter produces a bend in the DNA and increases the sensitivity to DNase I attack of nucleotides at positions −95, −83, −73, −62 and −56 in the early strand and at −102, −97, −85, −75 and −64 in the late strand (2,14; also shown in Fig. 5). In the absence of protein p4, RNA polymerase does not bind to the A3 promoter, but rather to the early A2b promoter, which is reflected by the appearance of a hypersensitive band at position −80 in the early strand (17; see also Fig. 5). The A2b promoter is known to be strong *in vivo* but weak in our *in vitro* transcription system (15). When protein p4 is present, the RNA polymerase is displaced from the early A2b promoter and instead it binds to the overlapping late A3 promoter (17; F. Rojo and M. Salas, submitted; see also Fig. 5). The binding of protein p4 to $P_{A3}\Delta 95$ produced a similar protection pattern as in the wild-type template, both in the absence or presence of RNA polymerase, and stimulated the binding of the RNA polymerase to the A3 promoter (Fig. 5). When $P_{A3}\Delta 87$ and $P_{A3}\Delta 88$ deletion derivatives were analyzed, the protein p4 protection pattern was very weak, consistent with an impaired binding of the activator to these DNAs in which protein p4 recognition sequences have been modified. Interestingly, in the presence of RNA polymerase, binding of protein p4 to the promoter was partially restored, and both protein p4 and RNA polymerase were able to bind to the P_{A3} mutants, although less efficiently than to the wild-type P_{A3}. This is consistent with the partial activity of these mutant promoters in the *in vitro* transcription assays. Similar results were obtained when the binding was analyzed by gel-retardation assays (not shown). It is worth noting that, although the DNase I footprint of the RNA polymerase in these latter mutant promoters was equivalent to that of the wild-type promoter, there were significant changes in the footprint pattern observed for protein p4. Indeed, the protein did not protect from the DNase I attack the vector sequences substituting the distal inverted sequence (centered at position −93), and the hypersensitive bands at positions −102 at the late strand and −95 at the early strand were missing. This observation indicates that, although protein p4 is recognizing the downstream part of its binding site, not affected by the deletion, it is not able to contact efficiently the non-specific sequences that substitute for the deleted region. The observed footprint corresponds indeed to only the downstream part of protein p4 binding region.

DISCUSSION

The current model for the binding of protein p4 to the phage ϕ29 late A3 promoter proposes that the activator binds as a dimer to two 8 bp long inverted sequences located around nt −67 and −93 relative to the transcription start site (2). This model was based on two experimental results: binding of protein p4 to the promoter protects from DNase I attack a region spanning nt −56 to −102, and methylation of guanine residues −67 and −93, located within these inverted sequences, drastically interfered with protein p4 binding to DNA. In this report we present evidence indicating that these inverted sequences play indeed a direct role in the recognition of P_{A3} by protein p4 . The missing nucleoside assay (7) has allowed us to evaluate the importance of each nucleoside of this DNA region in the binding process. The results indicate that the nucleosides whose removal was detrimental for protein p4 binding lied precisely in the above mentioned inverted repeats, which argues in favour of their involvement in protein p4 recognition of the promoter. When the sequences upstream of the inverted repeats were substituted by unrelated sequences, provided by the cloning vector (mutant $P_{A3}\Delta 95$), protein p4 binding and promoter activity were not impaired. Nevertheless, when the deletion affected the inverted sequence in the region located around nt −93 (mutants $P_{A3}\Delta 87$ and $P_{A3}\Delta 88$), protein p4 binding and promoter activity significantly decreased. It is noteworthy that the activity of the two latter mutants, although lower than that of the wild-type promoter, was higher than expected for the almost lack of binding of protein p4, as monitored by DNase I footprinting. However, when both protein p4 and RNA polymerase were simultaneously present, they helped each other to bind to the mutant promoters. The DNase I footprint for protein p4 bound to these promoter mutants in the presence of RNA polymerase suggested that the activator was in close contact with only part of its binding site, particularly with the region not affected by the deletion, that lies besides the RNA polymerase binding region. The distal side of this binding site, in which the recognition sequences had been partially eliminated and substituted by unrelated vector sequences, were not completely protected by protein p4 from DNase I attack, and the positions hypersensitive to DNase I digestion in the wild-type promoter were absent. This digestion pattern clearly showed that protein p4 was not stably bound to the modified region of the mutants mentioned above. Since the activator cannot bind to this altered recognition site by itself, it seems that RNA polymerase is helping it to bind to the promoter. This is best explained assuming that there are direct contacts between both proteins, that are responsible for the strong cooperativity observed in their binding to the promoter, and that stabilize the complex in the mutant promoters. It should be mentioned that when guanine residue −67, located in protein p4 recognition sequences, was modified by site-directed mutagenesis, binding of protein p4 to DNA was also impaired unless RNA polymerase was present (F. Rojo and M. Salas, submitted).

The proposed contacts between phage ϕ29 protein p4 and RNA polymerase are relevant for the understanding of the mechanism of transcription activation. Direct contacts between an activator protein and the RNA polymerase have also been proposed in the case of lambda CI repressor (4), the NtrC activator (20, 22) and for the CRP protein (6, 13, 19). It is of interest to compare protein p4 with CRP, since both activators induce a strong curvature in the DNA when binding to it. It has been shown that substitution of the CRP binding site in the *E. coli gal* promoter by a strong sequence-directed curvature, correctly aligned relative to the RNA polymerase binding site, significantly increases *in vivo* transcription from the CRP stimulated promoter, this effect being absolutely dependent on the orientation of the curvature (3). This led to the proposal that CRP activation mechanism relies to a significant extent on the curvature that it induces in the DNA, which should be enough to favour the correct positioning of the RNA polymerase and the formation of an open complex. According to this, the postulated contacts with the RNA polymerase would be an additional factor in the activation process. Although a similar approach has not yet been followed in the case of protein p4, it is clear that p4-induced curvature in the A3 promoter is required in the activation process, since the promoter remains inactive when protein p4 is not able to induce a full bending of the DNA (14). In addition, the data presented in this paper suggest the existence of direct contacts between protein p4 and RNA polymerase. This is in agreement

with recent findings suggesting that protein p4 needs to be correctly aligned relative to the RNA polymerase binding site to activate transcription efficiently. The introduction of four bp between the binding sites of both proteins impaired transcription from P_{A3} (17). In addition, it was shown that if these binding sites were separated by complete helical repeats, therefore maintaining their relative orientation, activation of transcription was only possible when the intervening sequences had a bent conformation. Both observations suggest that activation requires an interaction between protein p4 and RNA polymerase (17). Two possibilities are therefore conceivable: either both p4-induced DNA bending and p4-RNA polymerase contacts have an independent but additive effect in the activation mechanism, or the role of the p4-induced curvature is just to correctly orient the activator relative to the RNA polymerase so that adequate contacts can be made.

ACKNOWLEDGEMENTS

We are grateful to A.A.Travers and M.Serrano for helpful discussions, to J.M.Lázaro for purification of RNA polymerase and to L.Villar for technical assistance. This investigation was aided by research grants from the National Institutes of Health (5R01 GM27242-11) and Dirección General de Investigación Científica y Técnica (PB87 0323), and an institutional grant from Fundación Ramón Areces. B.N. was holder of a pre-doctoral F.P.I. fellowship from M.E.C., and F.R. and I.B. of post-doctoral fellowships from C.S.I.C.

REFERENCES

1. Adhya, S. and Garges, S. (1990). *J. Biol. Chem.* **265**, 10797–10800.
2. Barthelemy, I. and Salas, M. (1989) *J. Mol. Biol.* **218**, 235–232.
3. Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. and Buc, H. (1989). *EMBO J.* **8**, 4289–4296.
4. Bushman, F. D. and Ptasne, M. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 9353–9357.
5. Galas, D.J. and Schmitz (1978). *Nucl. Acids Res.* **5**, 3157–3170.
6. Gaston, K., Bell, A., Klob, A., Buc, H. and Busby, S. (1990). *Cell* **62**, 733–743.
7. Hayes, J. J. and Tullius, T. D. (1989). *Biochemistry* **28**, 9521–9527.
8. Inciarte, M. R., Lázaro, J. M., Salas, M. and Viñuela, E. (1976). *Virology* **74**, 314–323.
9. Maxam, A. M. and Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.
10. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). In Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Mellado, R. P., Barthelemy, I. and Salas, M. (1986). *Nucl. Acids Res.* 14, 4731–4741.
12. Parkhill, J. and Brown, N. L. (1990). *Nucl. Acids Res.* 18, 5157–5162.
13. Ren, Y. L., Garges, S., Adhya, S. and Krakow, J. S. (1988). *Proc. Natl. Acad. Sci. USA* 85, 4138–4142.
14. Rojo, F., Zaballos, A. and Salas, M. (1990). *J. Mol. Biol.* 211, 713–725.
15. Salas, M., Barthelemy, I. and Mellado, R. P. (1986). In *Bacillus* Molecular Genetics and Biotechnology Applications, Academic Press, pp. 395–409.
16. Schleif, R. (1988). *Science* **241**, 1182–1187.
17. Serrano, M., Barthelemy, I. and Salas, M. (1991). *J. Mol. Biol.*, in press.
18. Sogo, J. M., Inciarte, M. R., Corral, J., Viñuela, E. and Salas, M. 1979).*J. Mol. Biol.* 127, 411–436.
19. Straney, D. C., Straney, S. B. and Crothers, D. M. (1989). *J. Mol. Biol.* **206**, 41–57.
20. Su, W., Porter, S., Kustu, S. and Echols, H. (1990). *Proc. Natl. Acad. Sci. USA* 87, 5504–5508.
21. Travers, A. A. (1989). *Ann. Rev. Biochem.* 58, 427–452.
22. Wedel, A.; Weiss, D.; Popham, D. and Kustu, S. (1990). *Science* **248**, 486–490.
23. Yoshikawa, H. and Ito, J. (1982). *Gene* **17**, 323–335.